



(11) **EP 3 070 947 A1**

(12) **EUROPEAN PATENT APPLICATION**

(43) Date of publication:
21.09.2016 Bulletin 2016/38

(51) Int Cl.:
H04N 19/70 ^(2014.01) **H04N 19/107** ^(2014.01)
H04N 19/46 ^(2014.01) **H04N 19/23** ^(2014.01)

(21) Application number: **16159361.1**

(22) Date of filing: **09.03.2016**

(84) Designated Contracting States:
**AL AT BE BG CH CY CZ DE DK EE ES FI FR GB
GR HR HU IE IS IT LI LT LU LV MC MK MT NL NO
PL PT RO RS SE SI SK SM TR**
Designated Extension States:
BA ME
Designated Validation States:
MA MD

(30) Priority: **17.03.2015 CN 201510115730**

(71) Applicant: **Huawei Technologies Co., Ltd.
Longgang District
Shenzhen, Guangdong 518129 (CN)**

(72) Inventors:
• **AI, Changquan**
518129 Shenzhen Guangdong (CN)
• **ZHANG, Shaobo**
518129 Shenzhen Guangdong (CN)
• **YANG, Haitao**
518129 Shenzhen Guangdong (CN)

(74) Representative: **Maiwald Patentanwalts GmbH
Engineering
Elisenhof
Elisenstrasse 3
80335 München (DE)**

(54) **VIDEO FRAME DECODING METHOD AND APPARATUS**

(57) Embodiments of the present invention provide a video frame decoding method and apparatus. The video frame decoding method in the present invention includes: acquiring a target time point of a video; obtaining a frame number of a corresponding random access frame according to the target time point; obtaining, according to a prestored dependency set and the frame number of the random access frame, a frame number of a frame that can be independently decoded, where the dependency set includes at least one dependency, the dependency represents a correspondence between a frame

number of a random access frame and a frame number of a frame that can be independently decoded, and the frame that can be independently decoded is a frame on which decoding of the random access frame needs to depend; and obtaining, according to the frame number of the frame that can be independently decoded, video data of the frame that can be independently decoded, and decoding the random access frame according to the video data of the frame that can be independently decoded. In the embodiments of the present invention, efficient and fast video decoding is implemented.

EP 3 070 947 A1

Description**TECHNICAL FIELD**

5 **[0001]** Embodiments of the present invention relate to video technologies, and in particular, to a video frame decoding method and apparatus.

BACKGROUND

10 **[0002]** A video sequence is generally stored according to a specific frame structure. H.264 is used as an example, and a video frame structure of H.264 may be, for example, IPP...PP IPP...PPIPP...PP..., where an I frame (random access frame) is a frame that can be independently decoded, and a P frame (common frame) is a frame that cannot be independently decoded and decoding of the P frame needs to depend on an I frame before the P frame. When a video with such a frame structure is dragged to play or edited, an I frame corresponding to a target time point for dragging or editing needs to be found first.

15 **[0003]** Generally, in scenarios such as video surveillance and static observation and photography, a video background remains unchanged in a quite long time. Based on this characteristic, video encoding efficiency can be significantly improved by using an optimized encoding technology. After the optimized encoding technology is used, a frame structure of a video sequence changes accordingly, including a background frame, a random access frame, and a common frame, where only the background frame can be independently decoded, and decoding of both the random access frame and a common frame needs to depend on the corresponding background frame. When a video with such a frame structure is dragged to play or edited, a random access frame corresponding to a target time point for dragging or editing may be found first; however, with only the random access frame, decoding cannot be performed, and all frames located before the random access frame need to be searched for in a traversal manner according to frame header data of the random access frame, so as to determine a background frame on which decoding of the random access frame depends.

25 **[0004]** However, a traversal method is a searching method with extremely low efficiency, thereby severely reducing efficiency of decoding a video frame.

SUMMARY

30 **[0005]** Embodiments of the present invention provide a video frame decoding method and apparatus, so as to implement efficient and fast video decoding.

[0006] According to a first aspect, an embodiment of the present invention provides a video frame decoding method, including:

35 acquiring a target time point of a video;
obtaining a frame number of a corresponding random access frame according to the target time point;
obtaining, according to a prestored dependency set and the frame number of the random access frame, a frame number of a frame that can be independently decoded, where the dependency set includes at least one dependency,
40 the dependency represents a correspondence between a frame number of a random access frame and a frame number of a frame that can be independently decoded, and the frame that can be independently decoded is a frame on which decoding of the random access frame needs to depend; and
obtaining, according to the frame number of the frame that can be independently decoded, video data of the frame that can be independently decoded, and decoding the random access frame according to the video data of the
45 frame that can be independently decoded.

[0007] With reference to the first aspect, in a first possible implementation manner of the first aspect, the method further includes:

50 obtaining, according to the prestored dependency set and the frame number of the random access frame, the frame number of the frame that can be independently decoded, where the dependency set includes frame numbers of all frames, which can be independently decoded, of the video, and the obtained frame number of the frame that can be independently decoded is a frame number that is smaller than the frame number of the random access frame in the dependency set, where a difference between the frame number that can be independently decoded and the
55 frame number of the random access frame is the smallest.

[0008] With reference to the first aspect or the first possible implementation manner of the first aspect, in a second possible implementation manner of the first aspect, the dependency set is prestored in the video, or the dependency

set is prestored in an attached file of the video; and

before the obtaining, according to a prestored dependency set and the frame number of the random access frame, a frame number of a frame that can be independently decoded, the method further includes:

5 acquiring the dependency set from the video or the attached file.

[0009] With reference to the first aspect or either of the first and the second possible implementation manners of the first aspect, in a third possible implementation manner of the first aspect, the video includes video index information and video data, and the video index information includes: a correspondence between a time point of the video and a frame number of each frame of the video, frame numbers of all random access frames, and a correspondence between a frame number of a common frame in all the frames and a frame number of a random access frame; and the obtaining a frame number of a corresponding random access frame according to the target time point includes:

15 obtaining, according to the correspondence between a time point of the video and a frame number of each frame of the video, a frame number corresponding to the target time point; and
 if the frame number corresponding to the target time point is found in the frame numbers of all the random access frames, determining that the frame number of the random access frame corresponding to the target time point is the frame number corresponding to the target time point; or
 if the frame number corresponding to the target time point is not found in the frame numbers of all the random access frames, determining that the frame number corresponding to the target time point is a frame number of a common frame, and obtaining, according to the correspondence between a frame number of a common frame in all the frames and a frame number of a random access frame, the frame number of the random access frame corresponding to the target time point.

25 [0010] With reference to the third possible implementation manner of the first aspect, in a fourth possible implementation manner of the first aspect, the video index information further includes an index of video data corresponding to the frame number of each frame; and
 the obtaining, according to the frame number of the frame that can be independently decoded, video data of the frame that can be independently decoded includes:

30 querying the video index information according to the frame number of the frame that can be independently decoded, to obtain an index of the video data of the frame that can be independently decoded, and acquiring, according to the index, the video data of the frame that can be independently decoded.

35 [0011] According to a second aspect, an embodiment of the present invention provides a video frame decoding apparatus, including:

40 an acquiring module, configured to: acquire a target time point of a video, obtain a frame number of a corresponding random access frame according to the target time point, and obtain, according to a prestored dependency set and the frame number of the random access frame, a frame number of a frame that can be independently decoded, where the dependency set includes at least one dependency, the dependency represents a correspondence between a frame number of a random access frame and a frame number of a frame that can be independently decoded, and the frame that can be independently decoded is a frame on which decoding of the random access frame needs to depend; and
 45 a decoding module, configured to: obtain, according to the frame number of the frame that can be independently decoded, video data of the frame that can be independently decoded, and decode the random access frame according to the video data of the frame that can be independently decoded.

50 [0012] With reference to the second aspect, in a first possible implementation manner of the second aspect, the acquiring module is further configured to obtain, according to the prestored dependency set and the frame number of the random access frame, the frame number of the frame that can be independently decoded, where the dependency set includes frame numbers of all frames, which can be independently decoded, of the video, a quantity of all the frames that can be independently decoded is less than a quantity of all frames that are of the video and whose frame numbers are smaller than the frame number of the random access frame, and the obtained frame number of the frame that can be independently decoded is a frame number that is smaller than the frame number of the random access frame in the dependency set, where a difference between the frame number that can be independently decoded and the frame number of the random access frame is the smallest.

[0013] With reference to the second aspect or the first possible implementation manner of the second aspect, in a

second possible implementation manner of the second aspect, the dependency set is prestored in the video, or the dependency set is prestored in an attached file of the video; and the acquiring module is further configured to acquire the dependency set from the video or the attached file.

[0014] With reference to the second aspect or either of the first and the second possible implementation manners of the second aspect, in a third possible implementation manner of the second aspect, the video includes video index information and video data, and the video index information includes: a correspondence between a time point of the video and a frame number of each frame of the video, frame numbers of all random access frames, and a correspondence between a frame number of a common frame in all the frames and a frame number of a random access frame; and the acquiring module is specifically configured to: obtain, according to the correspondence between a time point of the video and a frame number of each frame of the video, a frame number corresponding to the target time point; and if the frame number corresponding to the target time point is found in the frame numbers of all the random access frames, determine that the frame number of the random access frame corresponding to the target time point is the frame number corresponding to the target time point; or if the frame number corresponding to the target time point is not found in the frame numbers of all the random access frames, determine that the frame number corresponding to the target time point is a frame number of a common frame, and obtain, according to the correspondence between a frame number of a common frame in all the frames and a frame number of a random access frame, the frame number of the random access frame corresponding to the target time point.

[0015] With reference to the third possible implementation manner of the second aspect, in a fourth possible implementation manner of the second aspect, the video index information further includes an index of video data corresponding to the frame number of each frame; and the decoding module is specifically configured to: query the video index information according to the frame number of the frame that can be independently decoded, to obtain an index of the video data of the frame that can be independently decoded, and acquire, according to the index, the video data of the frame that can be independently decoded.

[0016] According to the video frame decoding method and apparatus in the embodiments of the present invention, a dependency set is prestored, a frame number of a frame that can be independently decoded and on which decoding of a random access frame corresponding to a target time point needs to depend can be directly acquired according to the dependency set, and the random access frame is decoded by using video data of the frame that can be independently decoded, which implements efficient and fast video decoding.

BRIEF DESCRIPTION OF DRAWINGS

[0017] To describe the technical solutions in the embodiments of the present invention or in the prior art more clearly, the following briefly introduces the accompanying drawings required for describing the embodiments or the prior art. Apparently, the accompanying drawings in the following description show some embodiments of the present invention, and persons of ordinary skill in the art may still derive other drawings from these accompanying drawings without creative efforts.

FIG. 1 is a schematic diagram of a storage structure of a video file in the prior art;
 FIG. 2 is a schematic diagram of a storage structure of a video frame attribute in the prior art;
 FIG. 3 is a flowchart of a video frame decoding method according to an embodiment of the present invention;
 FIG. 4 is a schematic diagram of a storage structure of a video frame attribute according to the present invention;
 FIG. 5 is a schematic diagram of a video frame structure; and
 FIG. 6 is a schematic structural diagram of a video frame decoding apparatus according to an embodiment of the present invention.

DESCRIPTION OF EMBODIMENTS

[0018] To make the objectives, technical solutions, and advantages of the embodiments of the present invention clearer, the following clearly and completely describes the technical solutions in the embodiments of the present invention with reference to the accompanying drawings in the embodiments of the present invention. Apparently, the described embodiments are some but not all of the embodiments of the present invention. All other embodiments obtained by persons of ordinary skill in the art based on the embodiments of the present invention without creative efforts shall fall within the protection scope of the present invention.

[0019] The International Organization for Standardization (International Organization for Standardization, ISO for short)/International Electrotechnical Commission (International Electrotechnical Commission, IEC for short) 14496-12 standard provides a method for storing video data such as audio and a video. FIG. 1 is a schematic diagram of a storage structure of a video file in the prior art. As shown in FIG. 1, a video file mainly includes two parts: a video frame attribute (moov) and video data (mdat), and such a storage structure is generally referred to as a box (box), for example, a moov

box and an mdat box. When a video is dragged to play, edited, or the like, an index of an I frame corresponding to a target time point of the video needs to be found in the moov box first, and then, video data of the I frame is acquired from the mdat box.

[0020] FIG. 2 is a schematic diagram of a storage structure of a video frame attribute in the prior art. As shown in FIG. 2, a moov box may further include multiple boxes, where a box related to dragging a video to play, editing a video, or the like is an stbl box that stores index information of the video, and the box includes: an stts box that stores a correspondence between a time point of the video and a frame number of each frame of the video, an stsc box that stores a correspondence between a frame number and a video chunk (chunk) in which the frame number is located, an stsz box that stores a frame size, an stco box that stores an offset of each chunk, and an stss box that stores a frame number of a random access frame and a correspondence between a frame number of a common frame in all frames of the video and a frame number of a random access frame. A process of dragging a video to play, editing a video, or the like includes: searching the stts box according to a target time point of the video to find a frame number J corresponding to the target time point; if the stss box exists, searching the stss box for a frame number K of a random access frame corresponding to the frame number J, or if the stss box does not exist, determining that $K = J$; searching, according to the frame number K, the stsc box for a video chunk (chunk) L in which the random access frame is located; searching the stco box for a location offset M of the chunk L; calculating, according to the frame number K and the stsz box, an offset N that is of the random access frame and is in the chunk L; and acquiring video data of the random access frame from the mdat box according to M and N. Generally, the video starts to be played from a time point of the video; if the time point is exactly corresponding to a random access frame, the video directly starts to be decoded and played from the random access frame after video data of the random access frame is acquired; if the time point is corresponding to a common frame, the video also starts to be played from a random access frame after video data of the random access frame on which decoding of the common frame needs to depend is acquired. Dragging the video to play, editing the video, or the like is completed by means of cooperation of the foregoing boxes.

[0021] It may be learned from the foregoing process that, for dragging a conventional video to play, editing a conventional video, or the like, a common frame that depends on a random access frame can be decoded only after the random access frame is found. However, for videos that have a large quantity of identical or similar background areas, a video sequence obtained after the videos includes three types of frames: a background frame, a random access frame, and a common frame are encoded, where only the background frame can be independently decoded, and decoding of both the random access frame and the common frame needs to depend on the corresponding background frame. When a video with such a frame structure is dragged to play or edited, a random access frame corresponding to a target time point for dragging or editing may be found first; however, with only the random access frame, decoding cannot be performed, and all frames located before the random access frame need to be searched for one by one in a traversal manner according to frame header data of the random access frame, so as to determine a background frame on which decoding of the random access frame depends. However, a traversal method is a searching method with extremely low efficiency, thereby severely reducing efficiency of decoding a video frame.

[0022] FIG. 3 is a flowchart of a video frame decoding method according to an embodiment of the present invention. As shown in FIG. 3, the method in this embodiment may include the following steps:

Step 101: Acquire a target time point of a video.

[0023] This embodiment may be executed by a video frame decoding apparatus, and the apparatus may be a decoding module of any video player, or may be an independent decoding device. A video sequence generated after optimized encoding is performed on videos that have a large quantity of identical or similar background areas includes three types of frames: a background frame, a random access frame, and a common frame, where the background frame can be independently decoded, and decoding of both the random access frame and the common frame needs to depend on the background frame. In this embodiment and subsequent embodiments, the background frame is referred to as a frame that can be independently decoded, which is represented by RAP_I, and the random access frame is represented by RAP_P. When dragging a video to play, editing a video, or the like, a user generally drags a play slider of a player to select a target time point, and the video frame decoding apparatus may acquire the target time point of the video by using a location of the play slider.

[0024] Step 102: Obtain a frame number of a corresponding random access frame according to the target time point.

[0025] FIG. 4 is a schematic diagram of a storage structure of a video frame attribute according to the present invention. As shown in FIG. 4, functions of an stts box, an stsc box, an stsz box, an stco box, and an stss box are the same as those shown in FIG. 2, and details are not described herein again. A difference between FIG. 4 and FIG. 2 lies in that a samd box is added, where the samd box is used to store a dependency set in the present invention. The dependency set is described in detail in step 103. The video in this embodiment may use the storage structure shown in FIG. 4. A correspondence between a time point of the video and a frame number of each frame of the video is recorded in the video; a frame number corresponding to the target time point may be acquired by using the correspondence; then, it

may be determined, according to a frame number, recorded in the video, of a RAP_P frame, whether the frame number corresponding to the target time point is the frame number of the RAP_P frame or a frame number of a common frame; if the frame number is corresponding to the RAP_P frame, the frame number is the frame number of the RAP_P frame corresponding to the target time point, or if the frame number is corresponding to a common frame, a frame number of a RAP_P frame corresponding to the frame number that is obtained according to a correspondence, stored in the video, between a frame number of a common frame in all frames of the video and a frame number of a RAP_P frame is the frame number of the RAP_P frame corresponding to the target time point.

[0026] Step 103: Obtain, according to a prestored dependency set and the frame number of the random access frame, a frame number of a frame that can be independently decoded, where the dependency set includes at least one dependency, the dependency represents a correspondence between a frame number of a random access frame and a frame number of a frame that can be independently decoded, and the frame that can be independently decoded is a frame that can be independently decoded in the video.

[0027] The dependency set may be prestored, as shown in FIG. 4, in the samd box newly added into the video, or the dependency set may be prestored in an attached file of the video. It should be noted that, in addition to the foregoing two manners, the dependency may be prestored by using another storage method commonly used by persons skilled in the art, which is not specifically limited in the present invention. When the video is encoded, three types of frames, namely, a RAP_I frame, a RAP_P frame, and a common frame are generated, and frames of the video are numbered in an encoding order. Therefore, accordingly, in an encoding manner, both a RAP_P frame on which decoding of a common frame depends during decoding and a RAP_I frame on which decoding of the RAP_P frame depends during decoding are determined during encoding. For dragging a video to play, editing a video, or the like, generally, the video starts to be played from a time point of the video; if the time point is exactly corresponding to a RAP_P frame, the video directly starts to be decoded and played from the RAP_P frame after video data of the RAP_P frame is acquired; if the time point is corresponding to a common frame, the video also starts to be played from a RAP_P frame after video data of the RAP_P frame on which decoding of the common frame needs to depend is acquired. However, decoding of the video data of the RAP_P frame also needs to depend on video data of a RAP_I frame, where the video data of the RAP_I frame may be obtained by using the foregoing prestored dependency set. After the frame number of the RAP_P frame is obtained by performing step 102, the needed frame number of the RAP_I frame is obtained according to a correspondence, represented by a dependency in the dependency set, between a frame number of a RAP_P frame and a frame number of a RAP_I frame.

[0028] Optionally, the dependency set may also be a set of frame numbers of all RAP_I frames of the video, and the needed frame number of the RAP_I frame may be acquired from the set according to a preset rule. Specifically, all frame numbers that are smaller than a frame number of a to-be-decoded RAP_P frame in the dependency set are found first; then, a frame number, a difference between which and the frame number of the to-be-decoded RAP_P frame is smallest, is found in the frame numbers, and the frame number is the needed frame number of the RAP_I frame.

[0029] Step 104: Obtain, according to the frame number of the frame that can be independently decoded, video data of the frame that can be independently decoded, and decode the random access frame according to the video data of the frame that can be independently decoded.

[0030] The video further includes an index of video data corresponding to the frame number of each frame, for example, in the mdat box that stores video data, a storage address is used to represent an index of the video data; therefore, an address offset of a corresponding frame may be acquired by using a frame number, a storage location of the video data may be obtained by using the offset, and further the video data is read. The video data in this embodiment is frame data generated after the video is encoded, and specific content is related to a used encoded manner. A reason why a RAP_I frame can be independently decoded is that complete image data, including a pixel, an area, and the like, of the frame is recorded after the frame is encoded. A reason why a RAP_P frame cannot be independently decoded is that encoding of the frame depends on a RAP_I frame and is incremental encoding that is based on the RAP_I frame; therefore, only incremental data of the frame relative to the RAP_I frame is recorded, and the RAP_P frame can be decoded only after video data of the RAP_I frame and the incremental data of the RAP_P frame are acquired.

[0031] According to this embodiment, a dependency set is prestored, a frame number of a frame that can be independently decoded and on which decoding of a random access frame corresponding to a target time point needs to depend can be directly acquired according to the dependency set, and the random access frame is decoded by using video data of the frame that can be independently decoded, which implements efficient and fast video decoding.

[0032] The following describes in detail the technical solution of the method embodiment shown in FIG. 3 by using several specific embodiments.

Embodiment 1

[0033] A video uses the storage structure shown in FIG. 4, and a samd box is added into an stbl box, where the samd box stores a dependency set. The dependency set in this embodiment includes a frame number of a RAP_I frame in

each dependency and a frame number of a RAP_P frame in each dependency. Further, the dependency set may include: an identifier of the video, a quantity of dependencies, a quantity of RAP_I frames in each dependency, and a quantity of RAP_P frames in each dependency.

[0034] A specific storage structure of the samd box (that is, SampleDependencyBox) is as follows (for a uniform description, the present invention uses a description method provided in the ISO/IEC 14496-12 standard):

```

    aligned(8) class SampleDependencyBox
    extends FullBox('samd', version = 0, 0){
    unsigned int(32) track_ID;
    unsigned int(32) entry_count;
    int i;
    for(i = 0; i < entry_count; i++){
        dependencyEntryBox();
    }
    aligned(8) class dependencyEntryBox
    extends FullBox('deet', version = 0, 0){
    unsigned int(32) SampleSize1;
    unsigned int(32) SampleSize2;
    for(i = 0; i < SampleSize1; i++) {
    unsigned int(32) sample_num_1;
    }
    for(i = 0; i < SampleSize2; i++){
    unsigned int(32) sample_num_2;
    }
    }

```

[0035] Notes:

track_ID represents an identifier of a video;

entry_count represents a quantity of dependencies, where a dependency set includes at least one dependency, each dependency represents a correspondence between a frame number of a RAP_P frame and a frame number of a RAP_I frame, each dependency in this embodiment is described by using one structure body, and the structure body is described by using a deet box (that is, dependencyEntryBox);

SampleSize1 represents a quantity of RAP_I frames in each dependency;

SampleSize2 represents a quantity of RAP_P frames in each dependency;

sample_num_1 represents a frame number of a RAP_I frame in each dependency; and

sample_num_2 represents a frame number of a RAP_P frame in each dependency.

[0036] For example, FIG. 5 is a schematic diagram of a video frame structure. As shown in FIG. 5, there are two RAP_I frames whose frame numbers are 1 and 20, and there are six RAP_P frames whose frame numbers are 2, 8, 14, 21, 27, and 33. Decoding of the RAP_P frames whose frame numbers are 2, 8, and 14 depends on the RAP_I frame whose frame number is 1, and decoding of the RAP_P frames whose frame numbers are 21, 27, and 33 depends on the RAP_I frame whose frame number is 20. According to the foregoing description method, the dependency set in this embodiment is:

```

samd
track_ID = 1 (the identifier of the video)
entry_count = 2 (the quantity of dependencies)
deet (the first dependency)
SampleSize1 = 1 (the quantity of RAP_I frames)
SampleSize2 = 3 (the quantity of RAP_P frames)
[1] (the frame number of the RAP_I frame)
[2, 8, 14] (the frame numbers of the RAP_P frames)
deet (the second dependency)
SampleSize1 = 1 (the quantity of RAP_I frames)
SampleSize2 = 3 (the quantity of RAP_P frames)
[20] (the frame number of the RAP_I frame)
[21, 27, 33] (the frame numbers of the RAP_P frames)

```

[0037] Based on the foregoing storage structure, according to the samd box, a video frame decoding apparatus in this embodiment finds a frame number of a RAP_P frame, and determines whether the frame number belongs to the first dependency or the second dependency; a frame number of a RAP_I frame in a dependency to which the frame number belongs is a frame number of a frame that can be independently decoded and on which decoding of the RAP_P frame needs to depend. Therefore, for dragging a video to play, editing a video, or the like, video frame decoding steps include:

- (1) Query an stts box according to a target time point of the video to find a frame number J corresponding to the target time point.
- (2) Search an stss box for a frame number K of a RAP_P frame corresponding to the frame number J; if the frame number J is corresponding to the RAP_P frame, K = J.
- (3) Obtain a frame number R of a RAP_I frame according to a dependency set stored in a samd box and the frame number K.
- (4) Acquire, according to an stsc box, a number L of a chunk in which the frame number R is located.
- (5) Acquire an offset M of the chunk L according to an stco box.
- (6) Calculate, according to an stsz box, an offset N that is of the frame number R and is in the chunk L.
- (7) Acquire video data of the RAP_I frame whose frame number is R from an mdat box according to M and N.
- (8) Then, acquire, according to the method of step (4) to step (7), incremental video data of the RAP_P frame whose frame number is K.
- (9) Decode the RAP_P frame with reference to the video data of the RAP_I frame and the incremental video data of the RAP_P frame.

Embodiment 2

[0038] A video still uses the storage structure shown in FIG. 4, and a samd box is added into an stbl box, where the samd box stores a dependency set. The dependency set in this embodiment includes frame numbers of RAP_I frames in all dependencies. Further, the dependency set may include an identifier of the video and a quantity of dependencies.

[0039] A specific storage structure of the samd box (that is, SampleDependencyBox) is as follows:

```
aligned(8) class SampleDependencyBox
    extends FullBox('samd', version = 0, 0){
    unsigned int(32) track_ID;
    unsigned int(32) entry_count;
    int i;
    for(i = 0; i < entry_count; i++) {
    unsigned int(32) sample_num;
    }
    }
```

[0040] Notes:

track_ID represents an identifier of a video;
entry_count represents a quantity of dependencies; and
sample_num represents frame numbers of RAP_I frames in all dependencies.

[0041] The video frame structure shown in FIG. 5 is still used as an example, and the dependency set in this embodiment is:

```
samd
track_ID = 1 (the identifier of the video)
entry_count = 2 (the quantity of dependencies)
[1, 20] (the frame numbers of the RAP_I frames in all the dependencies)
```

[0042] Based on the foregoing storage structure, according to the samd box, a video frame decoding apparatus in this embodiment finds a frame number of a RAP_I frame on which decoding of a RAP_P frame needs to depend. The frame number of the RAP_I frame is a frame number that is smaller than a frame number of the RAP_P frame in the dependency set, where a difference between the frame number of the RAP_I frame and the frame number of the RAP_P frame is the smallest. In this embodiment, frame numbers of RAP_P frames are 2, 8, 14, 21, 27, and 33, and frame

numbers of RAP_I frames are 1 and 20. According to the foregoing matching rule, a RAP_I frame whose frame number is smaller than 2, 8, and 14 is only the RAP_I frame whose frame number is 1, and therefore, decoding of the RAP_P frames whose frame numbers are 2, 8, and 14 depends on the RAP_I frame whose frame number is 1. RAP_I frames whose frame numbers are smaller than 21, 27, and 33 are the RAP_I frames whose frame numbers are 1 and 20; however, differences between 20 and 21, 27, and 33 are smaller; therefore, decoding of the RAP_P frames whose frame numbers are 21, 27, and 33 depends on the RAP_I frame whose frame number is 20.

[0043] Video frame decoding steps in this embodiment are the same as those in Embodiment 1, and only a storage structure of a dependency set prestored in a samd box and a method for obtaining a RAP_I frame are different.

Embodiment 3

[0044] A video uses the storage structure shown in FIG. 2. In this embodiment, an stss box is extended, and a storage structure of a dependency set is added to an original storage structure, for example, a storage structure of an original stss box (that is, SyncSampleBox) is:

```
aligned(8) class SyncSampleBox
    extends FullBox('stss', version = 0, 0){
    unsigned int(32) entry_count;
    int i;
    for(i = 0; i < entry_count; i++) {
    unsigned int(32) sample_number;
    }
    }
```

[0045] A storage structure of an extended stss box (that is, SyncSampleBox) in this embodiment is:

```
aligned(8) class SyncSampleBox
    extends FullBox('stss', version = 1, 0){
    unsigned int(32) entry_count;
    int i;
    for(i = 0; i < entry_count; i++) {
    unsigned int(32) sample_number;
    }
    unsigned int(32) entry_count_1;
    int j;
    for(j = 0; j < entry_count_1; j++){
    unsigned int(32) sample_num_1;
    }
    }
```

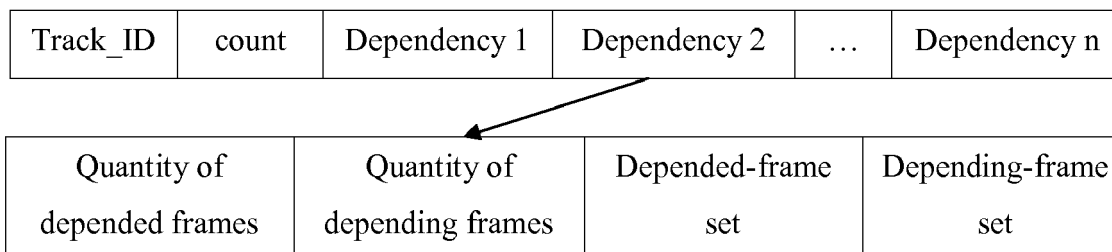
[0046] The storage structure of the dependency set in Embodiment 2 is added to the end of the storage structure of the original stss box. In this embodiment, for dragging a video to play, editing a video, or the like, video frame decoding steps include:

- (1) Query an stts box according to a target time point of the video to find a frame number J corresponding to the target time point.
- (2) Search an stss box for a frame number K of a RAP_P frame corresponding to the frame number J; if the frame number J is corresponding to the RAP_P frame, K = J.
- (3) Obtain a frame number R of a RAP_I frame according to a dependency set stored in an stss box and the frame number K.
- (4) Acquire, according to an stsc box, a number L of a chunk in which the frame number R is located.
- (5) Acquire an offset M of the chunk L according to an stco box.
- (6) Calculate, according to an stsz box, an offset N that is of the frame number R and is in the chunk L.
- (7) Acquire video data of the RAP_I frame whose frame number is R from an mdat box according to M and N.
- (8) Then, acquire, according to the method of step (4) to step (7), incremental video data of the RAP_P frame whose frame number is K.
- (9) Decode the RAP_P frame with reference to the video data of the RAP_I frame and the incremental video data of the RAP_P frame.

Embodiment 4

[0047] In this embodiment, a dependency set is separately stored, and the dependency set is stored in an attached file of a video, for example, the dependency set is stored in an a.index file. A storage structure that is of the dependency set and is in the attached file in this embodiment may be:

Track ID



[0048] Notes:

Track_ID represents an identifier of a video;
 count represents a quantity of dependencies; and
 dependency n represents an index of the n^{th} dependency.

[0049] Each dependency is represented by using the following quadruple:

quantity of depended frames, representing a quantity of RAP_I frames in each dependency;
 quantity of depending frames, representing a quantity of RAP_P frames in each dependency;
 depended-frame set, representing a frame number of a RAP_I frame in each dependency; and
 depending-frame set, representing a frame number of a RAP_P frame in each dependency.

[0050] Optionally, the storage structure that is of the dependency set and is in the attached file in this embodiment may be:

Track_ID	count	num1	num2	...	num n
----------	-------	------	------	-----	-------

[0051] Notes:

Track_ID represents an identifier of a video;
 count represents a quantity of dependencies; and
 num n represents a frame number of a RAP_I frame in the n^{th} dependency.

[0052] It is assumed that the video is a.mp4, and an attached file generated by using either of the foregoing storage structures is a.index. In this embodiment, for dragging a video to play, editing a video, or the like, video frame decoding steps include:

- (1) Query an stts box according to a target time point of the video to find a frame number J corresponding to the target time point.
- (2) Search an stss box for a frame number K of a RAP_P frame corresponding to the frame number J; if the frame number J is corresponding to the RAP_P frame, $K = J$.
- (3) Obtain a frame number R of a RAP_I frame according to a dependency set stored in a.index and the frame number K.
- (4) Acquire, according to an stsc box, a number L of a chunk in which the frame number R is located.
- (5) Acquire an offset M of the chunk L according to an stco box.
- (6) Calculate, according to an stsz box, an offset N that is of the frame number R and is in the chunk L.
- (7) Acquire video data of the RAP_I frame whose frame number is R from an mdat box according to M and N.

(8) Then, acquire, according to the method of step (4) to step (7), incremental video data of the RAP_P frame whose frame number is K.

(9) Decode the RAP_P frame with reference to the video data of the RAP_I frame and the incremental video data of the RAP_P frame.

[0053] According to the foregoing method embodiments, a dependency set is prestored, a frame number of a frame that can be independently decoded and on which decoding of a random access frame corresponding to a target time point needs to depend can be directly acquired according to the dependency set, and the random access frame is decoded by using video data of the frame that can be independently decoded, which implements efficient and fast video decoding.

[0054] FIG. 6 is a schematic structural diagram of a video frame decoding apparatus according to an embodiment of the present invention. As shown in FIG. 6, the apparatus in this embodiment may include an acquiring module 11 and a decoding module 12. The acquiring module 11 is configured to: acquire a target time point of a video, obtain a frame number of a corresponding random access frame according to the target time point, and obtain, according to a prestored dependency set and the frame number of the random access frame, a frame number of a frame that can be independently decoded, where the dependency set includes at least one dependency, the dependency represents a correspondence between a frame number of a random access frame and a frame number of a frame that can be independently decoded, and the frame that can be independently decoded is a frame on which decoding of the random access frame needs to depend. The decoding module 12 is configured to: obtain, according to the frame number of the frame that can be independently decoded, video data of the frame that can be independently decoded, and decode the random access frame according to the video data of the frame that can be independently decoded.

[0055] The apparatus in this embodiment may be configured to execute the technical solution in the method embodiment shown in FIG. 3. Implementation principles and technical effects in this embodiment are similar to those in the method embodiment, and details are not described herein again.

[0056] Further, the acquiring module 11 is further configured to obtain, according to the prestored dependency set and the frame number of the random access frame, the frame number of the frame that can be independently decoded, where the dependency set includes frame numbers of all frames, which can be independently decoded, of the video, and the obtained frame number of the frame that can be independently decoded is a frame number that is smaller than the frame number of the random access frame in the dependency set, where a difference between the frame number that can be independently decoded and the frame number of the random access frame is the smallest.

[0057] Further, the dependency set is prestored in the video, or the dependency set is prestored in an attached file of the video.

[0058] The acquiring module 11 is further configured to acquire the dependency set from the video or the attached file.

[0059] Further, the video includes video index information and video data, and the video index information includes: a correspondence between a time point of the video and a frame number of each frame of the video, frame numbers of all random access frames, and a correspondence between a frame number of a common frame in all the frames and a frame number of a random access frame.

[0060] The acquiring module 11 is specifically configured to: obtain, according to the correspondence between a time point of the video and a frame number of each frame of the video, a frame number corresponding to the target time point; and if the frame number corresponding to the target time point is found in the frame numbers of all the random access frames, determine that the frame number of the random access frame corresponding to the target time point is the frame number corresponding to the target time point; or if the frame number corresponding to the target time point is not found in the frame numbers of all the random access frames, determine that the frame number corresponding to the target time point is a frame number of a common frame, and obtain, according to the correspondence between a frame number of a common frame in all the frames and a frame number of a random access frame, the frame number of the random access frame corresponding to the target time point.

[0061] Further, the video index information further includes an index of video data corresponding to the frame number of each frame.

[0062] The decoding module 12 is specifically configured to: query the video index information according to the frame number of the frame that can be independently decoded, to obtain an index of the video data of the frame that can be independently decoded, and acquire, according to the index, the video data of the frame that can be independently decoded.

[0063] Persons of ordinary skill in the art may understand that all or some of the steps of the method embodiments may be implemented by a program instructing related hardware. The program may be stored in a computer-readable storage medium. When the program runs, the steps of the method embodiments are performed. The foregoing storage medium includes: any medium that can store program code, such as a ROM, a RAM, a magnetic disk, or an optical disc.

[0064] Finally, it should be noted that the foregoing embodiments are merely intended for describing the technical solutions of the present invention, but not for limiting the present invention. Although the present invention is described

in detail with reference to the foregoing embodiments, persons of ordinary skill in the art should understand that they may still make modifications to the technical solutions described in the foregoing embodiments or make equivalent replacements to some or all technical features thereof; however, these modifications or replacements do not make the essence of corresponding technical solutions depart from the scope of the technical solutions in the embodiments of the present invention.

Claims

1. A video frame decoding method, comprising:

acquiring a target time point of a video;
 obtaining a frame number of a corresponding random access frame according to the target time point;
 obtaining, according to a prestored dependency set and the frame number of the random access frame, a frame number of a frame that can be independently decoded, wherein the dependency set comprises at least one dependency, the dependency represents a correspondence between a frame number of a random access frame and a frame number of a frame that can be independently decoded, and the frame that can be independently decoded is a frame on which decoding of the random access frame needs to depend; and
 obtaining, according to the frame number of the frame that can be independently decoded, video data of the frame that can be independently decoded, and decoding the random access frame according to the video data of the frame that can be independently decoded.

2. The method according to claim 1, wherein the method comprises:

the obtained frame number of the frame that can be independently decoded is a frame number that is smaller than the frame number of the random access frame in the dependency set, wherein a difference between the frame number that can be independently decoded and the frame number of the random access frame is the smallest.

3. The method according to claim 1 or 2, wherein the dependency set is prestored in the video, or the dependency set is prestored in an attached file of the video; and before the obtaining, according to a prestored dependency set and the frame number of the random access frame, a frame number of a frame that can be independently decoded, the method further comprises:

acquiring the dependency set from the video or the attached file.

4. The method according to any one of claims 1 to 3, wherein the video comprises video index information and video data, and the video index information comprises: a correspondence between a time point of the video and a frame number of each frame of the video, frame numbers of all random access frames, and a correspondence between a frame number of a common frame in all the frames and a frame number of a random access frame; and the obtaining a frame number of a corresponding random access frame according to the target time point comprises:

obtaining, according to the correspondence between a time point of the video and a frame number of each frame of the video, a frame number corresponding to the target time point; and
 if the frame number corresponding to the target time point is found in the frame numbers of all the random access frames, determining that the frame number of the random access frame corresponding to the target time point is the frame number corresponding to the target time point; or
 if the frame number corresponding to the target time point is not found in the frame numbers of all the random access frames, determining that the frame number corresponding to the target time point is a frame number of a common frame, and obtaining, according to the correspondence between a frame number of a common frame in all the frames and a frame number of a random access frame, the frame number of the random access frame corresponding to the target time point.

5. The method according to claim 4, wherein the video index information further comprises an index of video data corresponding to the frame number of each frame; and the obtaining, according to the frame number of the frame that can be independently decoded, video data of the frame that can be independently decoded comprises:

querying the video index information according to the frame number of the frame that can be independently decoded, to obtain an index of the video data of the frame that can be independently decoded, and acquiring, according to the index, the video data of the frame that can be independently decoded.

5 **6.** A video frame decoding apparatus, comprising:

an acquiring module, configured to: acquire a target time point of a video, obtain a frame number of a corresponding random access frame according to the target time point, and obtain, according to a prestored dependency set and the frame number of the random access frame, a frame number of a frame that can be independently decoded, wherein the dependency set comprises at least one dependency, the dependency represents a correspondence between a frame number of a random access frame and a frame number of a frame that can be independently decoded, and the frame that can be independently decoded is a frame on which decoding of the random access frame needs to depend; and

a decoding module, configured to: obtain, according to the frame number of the frame that can be independently decoded, video data of the frame that can be independently decoded, and decode the random access frame according to the video data of the frame that can be independently decoded.

7. The apparatus according to claim 6, wherein the dependency set comprises frame numbers of all frames, which can be independently decoded, of the video, and the obtained frame number of the frame that can be independently decoded is a frame number that is smaller than the frame number of the random access frame in the dependency set, wherein a difference between the frame number that can be independently decoded and the frame number of the random access frame is the smallest.

8. The apparatus according to claim 6 or 7, wherein the dependency set is prestored in the video, or the dependency set is prestored in an attached file of the video; and the acquiring module is further configured to acquire the dependency set from the video or the attached file.

9. The apparatus according to any one of claims 6 to 8, wherein the video comprises video index information and video data, and the video index information comprises: a correspondence between a time point of the video and a frame number of each frame of the video, frame numbers of all random access frames, and a correspondence between a frame number of a common frame in all the frames and a frame number of a random access frame; and the acquiring module is specifically configured to: obtain, according to the correspondence between a time point of the video and a frame number of each frame of the video, a frame number corresponding to the target time point; and if the frame number corresponding to the target time point is found in the frame numbers of all the random access frames, determine that the frame number of the random access frame corresponding to the target time point is the frame number corresponding to the target time point; or if the frame number corresponding to the target time point is not found in the frame numbers of all the random access frames, determine that the frame number corresponding to the target time point is a frame number of a common frame, and obtain, according to the correspondence between a frame number of a common frame in all the frames and a frame number of a random access frame, the frame number of the random access frame corresponding to the target time point.

10. The apparatus according to claim 9, wherein the video index information further comprises an index of video data corresponding to the frame number of each frame; and the decoding module is specifically configured to: query the video index information according to the frame number of the frame that can be independently decoded, to obtain an index of the video data of the frame that can be independently decoded, and acquire, according to the index, the video data of the frame that can be independently decoded.

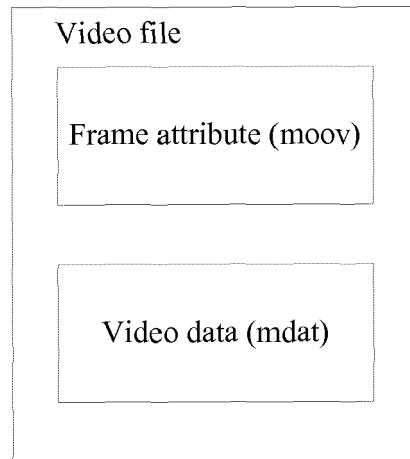


FIG. 1

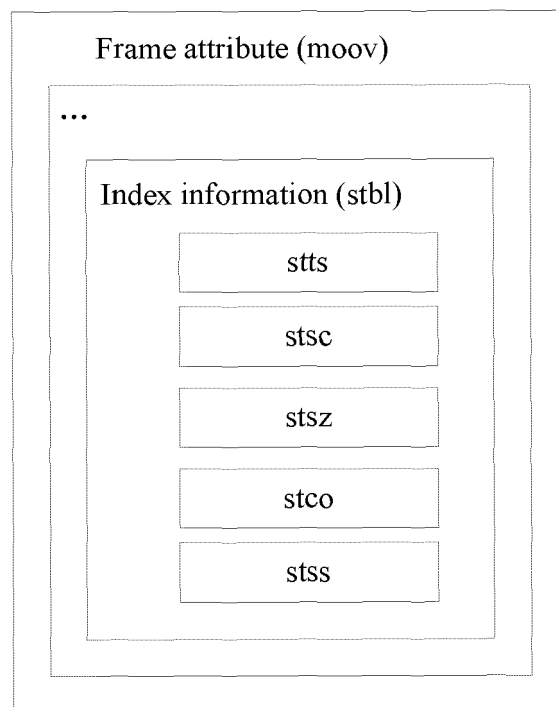


FIG. 2

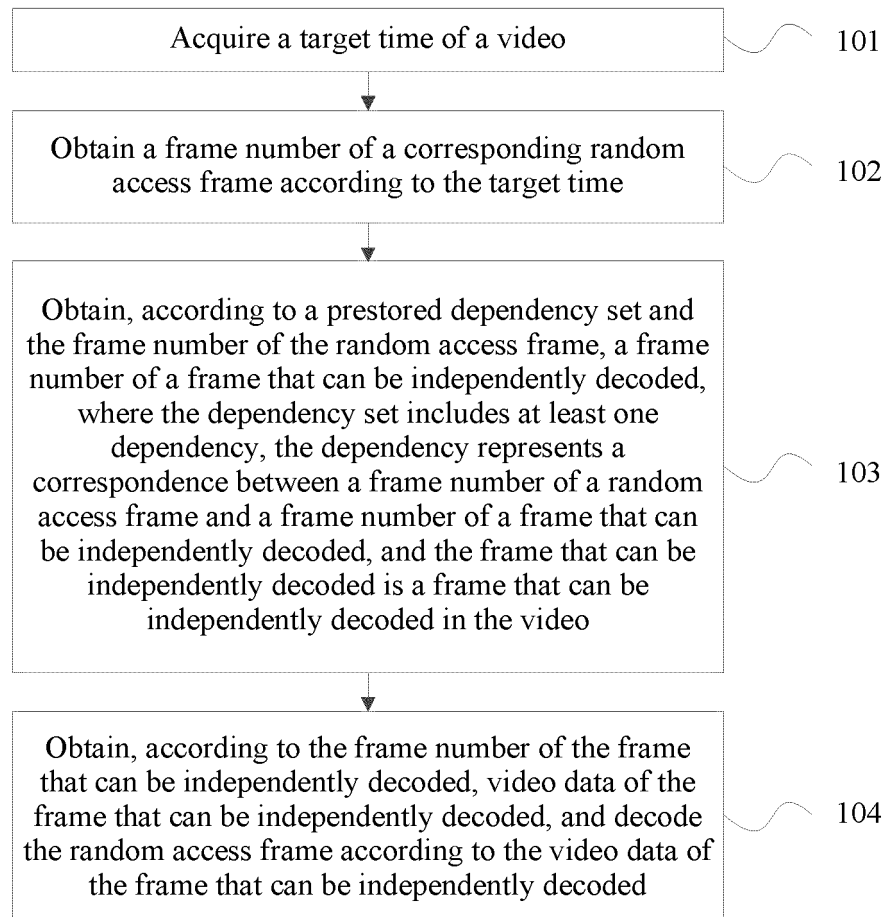


FIG. 3

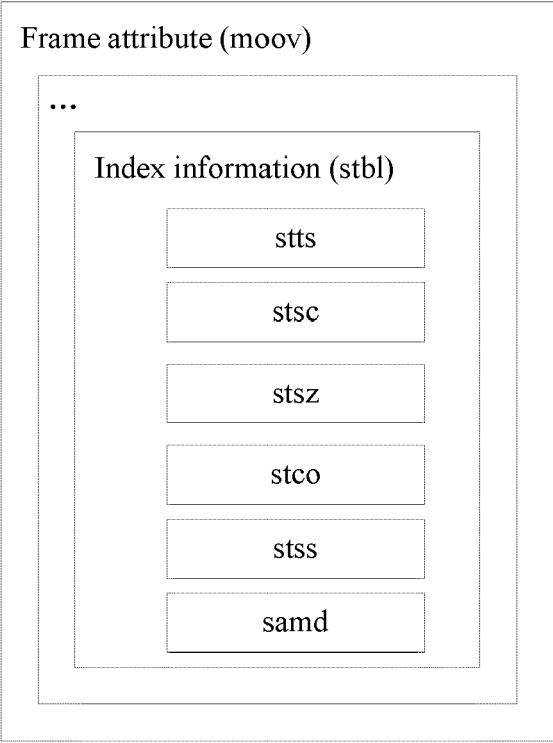


FIG. 4

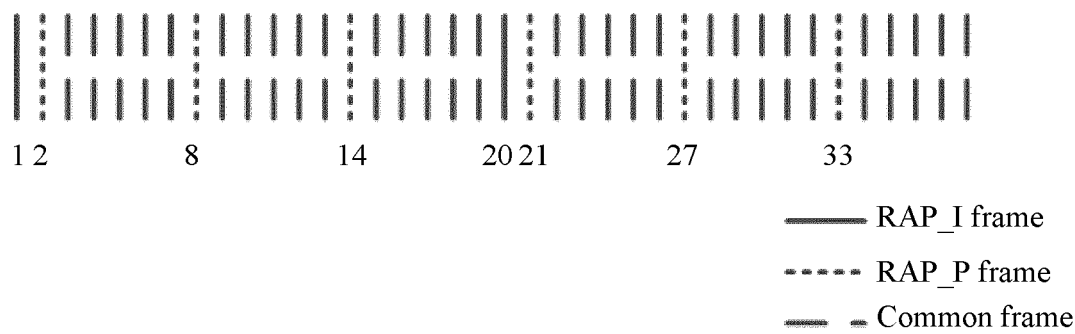


FIG. 5

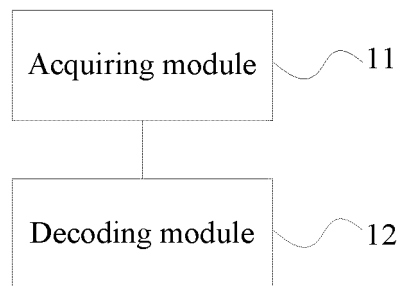


FIG. 6



EUROPEAN SEARCH REPORT

Application Number
EP 16 15 9361

5

10

15

20

25

30

35

40

45

50

55

DOCUMENTS CONSIDERED TO BE RELEVANT			
Category	Citation of document with indication, where appropriate, of relevant passages	Relevant to claim	CLASSIFICATION OF THE APPLICATION (IPC)
X	ANONYMOUS: "Text of ISO/IEC 14496-12 4th edition", 100. MPEG MEETING; 30-4-2012 - 4-5-2012; GENEVA; (MOTION PICTURE EXPERT GROUP OR ISO/IEC JTC1/SC29/WG11), no. N12640, 8 May 2012 (2012-05-08), XP030019114, * page 108 * * page 130 - page 132 * * page 26 - page 45 * -----	1-10	INV. H04N19/70 H04N19/107 H04N19/46 H04N19/23
X	DAVID SINGER: "Random Access in 14496-12, the ISO Base Media File Format", 96. MPEG MEETING; 21-3-2011 - 25-3-2011; GENEVA; (MOTION PICTURE EXPERT GROUP OR ISO/IEC JTC1/SC29/WG11), no. m20109, 16 March 2011 (2011-03-16), XP030048676, * the whole document *	1,6	
A	----- Y-K WANG ET AL: "File format - on RAP definition and signaling", 95. MPEG MEETING; 24-1-2011 - 28-1-2011; DAEGU; (MOTION PICTURE EXPERT GROUP OR ISO/IEC JTC1/SC29/WG11), no. m19132, 17 January 2011 (2011-01-17), XP030047699, * the whole document * -----	2-5,7-10	TECHNICAL FIELDS SEARCHED (IPC) H04N
A		1-10	
The present search report has been drawn up for all claims			
Place of search Munich		Date of completion of the search 27 May 2016	Examiner Cyranka, Oliver
CATEGORY OF CITED DOCUMENTS X : particularly relevant if taken alone Y : particularly relevant if combined with another document of the same category A : technological background O : non-written disclosure P : intermediate document T : theory or principle underlying the invention E : earlier patent document, but published on, or after the filing date D : document cited in the application L : document cited for other reasons & : member of the same patent family, corresponding document			

 1
EPO FORM 1503 03.82 (P04C01)